DOCUMENT RESUME

ED 246 063 TM 840 325

AUTHOR Olejnik, Stephen F.; Algina, James
TITLE A Review of Nonparametric Alternatives to Analysis of Covariance.
PUB DATE Apr 84
NOTE 46p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Covariance; Comparative Analysis; Hypothesis Testing; *Mathematical Formulas; *Nonparametric Statistics; Regression (Statistics); Statistical Analysis; Statistical Distributions
IDENTIFIERS *Power (Statistics); *Type I Errors; Violation of Assumptions

ABSTRACT

Five distribution-free alternatives to parametric analysis of covariance (ANCOVA) are presented and demonstrated using a specific data example. The procedures considered are those suggested by Quade (1967); Puri and Sen (1969); McSweeney and Porter (1971); Burnett and Barr (1978); and Shirley (1981). The results of simulation studies investigating these procedures regarding their respective Type I error rate under a null condition and their statistical power are also reviewed. The results indicate that the nonparametric procedures have appropriate Type I error rates only for those situations in which parametric ANCOVA is robust to violations of data assumptions. In terms of statistical power, nonparametric alternatives to parametric ANCOVA provide a considerable power advantage only for situations where extreme violations of assumptions have occurred and the linear relationship between measures is weak. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED246063

A Review of Nonparametric Alternatives to Analysis of Covaraince

Stephen F. Olejnik

James Algina

University of Florida

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. Olejnik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 840 325

Running head: Nonparametric ANCOVA

ABSTRACT

Five distribution-free alternatives to parametric analysis of covariance are presented and demonstrated using a specific data example. The results of simulation studies investigating these procedures regarding their respective Type I error rate under a null condition and their statistical power are also reviewed. The results indicate that the nonparametric procedures have appropriate Type I error rates only for those situations in which parametric ANCOVA is robust to violations of data assumptions. In terms of statistical power nonparametric alternatives to parametric ANCOVA provide a considerable power advantage only for situations where extreme violations of assumptions have occurred and the linear relationship between measures is weak.

A Review of Nonparametric Alternatives

to Analysis of Covariance

Nonparametric analysis strategies are often recommended in situations where population data have a non-normal distribution. In the case of the randomized pretest-posttest design (Campbell and Stanley, 1963) where analysis of covariance is generally the preferred parametric analysis strategy, it is the conditional distributions of the dependent variable which is assumed to be normal. Where conditional distributions are non-normal several distribution-free analysis procedures have been suggested as alternatives to analysis of covariance. It is the purpose of the present paper to discuss five of the more popular approaches and to review the results of monte-carlo studies which have investigated the Type I error rates and statistical power of these procedures under a variety of data conditions. The five procedures considered here are those suggested by 1) Quade (1967); 2) Puri and Sen (1969); 3) McSweeney and Porter (1971); Burnett and Barr (1978); and Shirley (1981).

In addition to conditional normality, parametric analysis of covariance assumes: 1) independence between experimental units; 2) a linear relationship between the pretest and posttest; 3) an equal relationship between the pretest and posttest across all groups; 4) equality of conditional variances within groups across all levels of the pretest; and 5) equality of conditional variances between groups. The distribution-free procedures considered here are less restrictive than parametric analysis of covariance in that conditional normality, a linear relationship and homogeneity of conditional variance within groups are not assumed. With respect to the linearity assumption, the

nonparametric procedures require only a monotonic increasing function relating the pretest and the posttest.

Over the years several analytic and monte-carlo studies have investigated the robustness of parametric analysis of covariance to violations of the data assumptions listed above. Glass, Peckham and Sanders (1972) and Elashoff (1969) have reviewed the early research findings and have concluded that the Type I error rates for analysis of covariance are not seriously affected by either 1) conditional non-normality or 2) small departures from slope equality if sample sizes are equal. More recently Hsu (1983) has summarized several research studies conducted since the publication of Glass et al.'s review. These investigations have focused on the effects of violating more than a single assumption in a data set. Hsu concluded that multiple violations can be a serious problem for analysis of covariance however with equal n the procedure is robust to heterogeneous regessions, and variances as well as non-normality.

Although parametric analysis of covariance appears to be robust to violations of certain data assumptions, the use of distribution-free analysis strategies are often recommended since research findings have indicated that nonparametric strategies can be more powerful than the parametric approach when data assumptions are violated (Blair and Higgins 1980). It has been the issue of statistical power which has been a focus of attention in the monte-carlo studies reviewed below on distribution-free alternatives to parametric analysis of covariance.

Before introducing each of the five nonparametric strategies, it is useful to briefly restate the procedure followed in parametric analysis of

covariance. To keep the presentation basic the analysis of a single factor model is considered with one covariate and one dependent variable. Furthermore the discussion is made concrete by analyzing a data set first reported by Quade (1967). The same data set is reanalyzed using each of the nonparametric approaches reviewed in the later section of the paper.

Parametric analysis of covariance. Perhaps the simplest way of thinking about analysis of covariance is through a regression context. From a regression perspective the effects of treatment conditions are evaluated through a comparison of intercepts estimated for separate regression lines for each treatment group assuming a constant within-group slope. The overall test for a treatment effective achieved through a comparison of 1) the squared multiple correlation for a regression model which contains both the pretest and the grouping factor as independent variables with 2) the multiple squared correlation for the regression model which includes the pretest data alone as an independent variable. The analysis of covariance can be completed in three steps:

1) Estimate the linear regression equation for the dependent variable on the covariate across all groups ($\hat{Y} = b_0 + b_{Y/X}X$) and computed the multiple squared correlation coefficient ($R^2_{Y/X}$).
2) Estimate the linear regression equation for the dependent variable on the covariate and grouping factor. If there are two groups, a single dummy coded variable (0,1) is needed to represent the grouping factor ($\hat{Y} = b_0 + b_{Y/X}X + b_{Y/G}G$). Compute the multiple squared correlation coefficient, $R^2_{Y/X,G}$.
3) Determine the R^2 increase associated with adding the grouping

factor to the regression model in step 1 and test for statistical significance with an F ratio: $$\frac{R^2_{Y/X,G} - R^2_{Y/X} / P_2 - P_1}{(1-R^2_{Y/X,G})/N-P_2}$$

Where P_1, P_2 are the number of parameters estimated in the regression models of steps 1 and 2 respectively;

N is the total sample size on which the parameters were estimated.

The computed F ratio can be compared with the theoretical F ratio under the null hypothesis with P_2-P_1, and $N-P_2$ degrees of freedom.

Table 1 contains a data set first reported by Quade (1967) and are used to demonstrate the three step procedure followed in the parametric analysis of covariance. Quade's data contained a total sample of 15 subjects with 5 individuals from each of three groups. Since the grouping factor contained three levels, two dummy coded variables were needed to represent the grouping factor.

Insert Table 1 Here

Step 1 resulted in the following regression equation:

$\hat{Y} = 32.6667 + 1.844X$ with $R^2_{Y/X} = .4926$.

Step 2 generated the regression equation:

$\hat{Y} = 33.599 + 2.021X - 22.959G_1 + 4.283G_2$ with $R^2_{Y/X,G_1G_2} = .5529$.

Step 3, the R^2 increase equalled .0603 and the computed F ratio calculated as

$$\frac{.5529 - .4926/4-2}{(1-.5529)/15-4}$$

$$= \frac{.0302}{.04065}$$

$$= .7429$$

The critical F ratio with 2 and 11 degrees of freedom at a .05 level of significance equals 3.98. The probability value associated with the computed F ratio equalled .4988. The results of this analysis therefore indicate that there is insufficient evidence to conclude that the three treatments had differential effects.

A test for the equality of the within group regression slopes can also be calculated by comparing the squared multiple correlation coefficient for the regression model which includes the pretest, the grouping factor and the interaction between the pretest and grouping factor as independent variables with the R^2 statistics computed in step 2. This comparison was made using Quade's data. With the interaction added to the model in step 2 the R^2 equalled .6498 and the F ratio for the R^2 increase associated with adding the interaction terms equalled 1.25 with a probability value of .3331. Thus there was insufficient evidence to indicate that the within group regression slopes differed across the three groups in the study.

Quade's distribution-free test. One of the first nonparametric alternatives to parametric analysis of covariance was developed by Quade (1967). Quade's procedure can be summarized in three steps.

1) Rank the dependent variable and the covariate separately across all groups.

2) Calculate the linear regression of the ranked dependent variable on the ranked covariate measure and calculate the residuals for the model.

3) Using the residuals calculated in step 2 as the dependent measure estimate the regression model with the grouping factor as the independent variable and test for significance the multiple squared correlation coefficient.

Applying these three steps to the data set reported in Table 1 provides the following results: Table 2 presents the ranked data from step 1. The regression

Insert Table 2 here

equation relating the ranked dependent variable on the ranked covariate was estimated as

$$\hat{RY} = 2.6 + .675\ RX.$$

The residuals for this model are reported in Table 3. Finally using the residuals

Insert Table 3 here

as the dependent variable the regression model relating this outcome to the grouping factor resulted in the following equation

$$\hat{E} = 1.205 - 3.08G_1 - .53G_2$$

with a squared multiple correlation coefficient equalling .1785. The computed F ratio for this model equalled 1.30 with a probability value of .3073.

Quade's procedure can be used for factorial designs and can include multiple covariate data. The approach assumes that the relationship between the pretest and the posttest is the same across all groups and does not provide for a procedure to test the assumption. Furthermore in contrast to parametric analysis of covariance where the computed test statistic considers both the within group and total group regression slopes, Quade's procedure

only uses the total group regression line. Finally in comparing Quade's approach with other nonparametric alternatives it is useful to note that the residuals calculated in step 2 of the procedure could also be calculated as an adjusted deviation score.

$$E = d_y - r_s d_x.$$

Where d_y is the deviation of an observed rank for the dependent variable from the grand mean rank across all groups, $(\frac{N+1}{2})$;

d_x is the deviation of the observed rank for the covariate from the grand mean of the ranked covariate;

r_s is the Spearman rank correlation coefficient across all groups.

Puri and Sen's solution. Two years following the publication of Quade's distribution-free alternative to analysis of covariance Puri and Sen (1969) developed and published a nonparametric alternative to ANCOVA which was more general than Quade's approach. Their solution could employ ranked data or any tranformation of the ranked data such as normal scores as long as the transformed scores were monotonically related to the original data. The procedure also provided an expression for determining the exact probability of a given set of data. With even moderate sample sizes however the calculations were prohibitive and the chi-square distribution was suggested as an appropriate reference distribution for their test statistic.

Puri and Sen's approach is a five step process which begins with

1) the ranking of each variable separately across all groups. Further transformation of the worked data is permitted if desired.

2) The within group means of the ranked variables ($\bar{r}_{y\cdot k}$, $\bar{r}_{x\cdot k}$) and the weighted grand rank mean ($\bar{r}..$) across all groups is calculated. The deviations of the group means from the grand mean is also computed ($d_{y\cdot k} = \bar{r}_{y\cdot k} - \bar{r}..$ and $d_{x\cdot k} = \bar{r}_{x\cdot k} - \bar{r}..$).

3) A residual mean within each group is computed by calculating the difference between the observed within group mean deviation for the dependent variable ($d_{y\cdot k}$) and the predicted mean deviation for the dependent variable based on the regression of the rank deviation scores for the posttest ($r_{y\cdot k}-\bar{r}..$) on the rank deviation scores for the pretest ($r_{x\cdot k}-\bar{r}..$) across all groups. With a single covariate these mean residuals can be calculated as:

$$E_{\cdot k} = d_{y\cdot k} - r_s d_{x\cdot k}.$$

Where r_s is the Spearman rank correlation for the total sample.

4) Order the ranked variables so the dependent variable is first and calculate the variance-covariance matrix (V) using N rather than N-1 in the denominator. Invert the matrix (V^{-1}) and isolate the first diagonal element (v^{11}). This element is the reciprocal of the error variance from the regression of individual transformed ranks. With a single covariate this element can be computed as:

$$v^{11} = 1/[\frac{N(N^2-1)}{12}(1-r_s^2)]$$

5. The test statistic is computed as:

$$L_n = v^{11}\sum_{k=1}^{K} n_k E_{\cdot k}^2$$

With large sample size L_n has a chi-square distribution with the number of groups minus one degrees of freedom (k-1).

Following Puri and Sen's procedure using the data in Table 2 the results are as follows:

Step 2, the within group means on the ranked posttest and pretest are 7.2, 9.6; 8.0, 7.0 and 8.8, 7.4 for groups 1, 2 and 3 respectively. Deviations of group rank means from the grand mean of the ranks were calculated as: -.8, 1.6; 0, -1 and .8, -.6 for the posttest and pretest in groups 1, 2 and 3 respectively.

Step 3, the Spearman rank correlation for the total data set equalled .675 and the within group residual means equalled -1.88, .675 and 1.205 for groups 1, 2 and 3 respectively.

Step 4, for the data in table 2 the variance-covariance matrix equalled:

$$V = \begin{vmatrix} 18.667 & 12.6 \\ 12.6 & 18.667 \end{vmatrix}$$

and the inverse was

$$V^{-1} \begin{vmatrix} .0984 & -.0664 \\ -.0664 & .0984 \end{vmatrix}$$

Step 5, the computed test statistic equalled

$$Ln = v^{11} \sum_{k=1}^{K} n_k E^2_{.k}$$

$$= .0984(27.21025)$$

$$= 2.6775$$

The critical Chi-square statistic with k-1 on 2 degrees of freedom equals 5.991 at the .05 level of significance.

The exact probability value for the computed test statistic was equal to .2622. The results therefore indicate that there is insufficient evidence to reject the null hypothesis that the distribution of the posttest scores differ across the three groups.

As with Quade's procedure, Puri and Sen's approach can be used in factorial designs and can include multiple covariates. It also assumes that the relationship between the pretest and posttest is constant across groups.

McSweeney and Porter's rank transformation. A third alternative to parametric analysis of covarinace, suggested by McSweeney and Porter (1971), differed from the two previous approaches by considering both the within group and total group regression slopes in computing the test statistic. Their approach recommended that the original data be transformed by replacing the original data with ranks across groups considering each variable separately. Using the ranked data the analysis proceeds following the same steps as parametric analysis of covariance. The rank transformation approach therefore has four steps:

1) Rank the dependent variable and the covariate separately across all groups.
2) Estimate the linear regression equation for the ranked dependent variable on the ranked covariate across all groups and compute the multiple squared correlation coefficient $(R^2_{RY/RX})$.
3) Estimate the linear regression equation for the ranked dependent variable on the ranked covariate and the grouping

factor and compute the multiple squared correlation coefficient ($R^2_{RY/RX,G}$).

4) Determine the R^2 increase associated with the grouping factor and test for statistical significance:

$$\frac{R^2_{RY/RX,G} - R^2_{RY/RX/P_2-P_1}}{(1-R^2_{RY/RX,G})/N-P_2}$$

Where P_1 is the number of parameters estimated in the reduced model computed in step 2

P_2 is the number of parameters estimated in the full model estimated in step 3.

The computed test statistic has an F distribution with P_2-P_1 and $N-P_2$ degrees of freedom.

Applying these steps with the data in table 1 provided the following results:

Step 1 is reported in table 2.

Step 2 resulted in the following equation $\hat{RY}=2.6+.675RX$ with $R^2_{RY/RX} = .4556$.

Step 3 resulted in generating the regression

$RY = 3.179 + .7596RX - 3.271G_1 - .4966_2$ with $R^2_{RY/RX_1G_1G_2} = .5595$.

Step 4. The computed test statistic equalled

$$\frac{.5595 - .4556/4-2}{(1-.5595)/15-4}$$

$$= \frac{.04195}{.0400}$$

$$= 1.30$$

The critical F statistic for 2 and 11 degrees of freedom equals 3.98. The probability value associated with the computed F statistic equalled .3122. There was insufficient evidence therefore to reject the null hypothesis that the distributions of the posttest scores differed across the three groups.

The rank transformation approach can also statistically test the equality of the within group relationship between the covariate and dependent variable across the three groups. This analysis is made by adding an interaction factor between the ranked covariate and the grouping factor to the regression model in step 3; computing the R^2 statistic for that model and comparing it with the R^2 from step 3. Using Quade's data the R^2 statistic for the interaction model equalled .6675 and the F ratio for the R^2 increase associated with the interaction terms equalled 1.46 with a probability value of .2821. The results therefore indicate that there was insufficient evidence to conclude the within group regression slopes differed across the treatment groups.

The rank transformation approach can be used in factorial designs and with multiple covariates.

Burnett and Barr's rank difference scores. Computationally the least complicated alternative to analysis of covariance was suggested by Burnett and Barr (1977). Their approach is a modification of Quade's solution in which it is assumed that the relationship between the ranks on posttest and pretest for the total sample is equal to one. With this assumption the residuals computed in Quade's second step are gain scores based on ranked

data. Burnett and Barr's solution can be summarized in three steps:

Step 1. For each variable rank the data across all groups.

Step 2. Calculate the difference between the ranked posttest and ranked pretest data.

Step 3. Determine the regression equation for the difference in ranks on the grouping factor and compute the multiple squared correlation coefficient. Test for statistical significance the multiple squared correlation coefficient.

Applying these steps to the data in table 1 provided the following results.

Steps 1 and 2 are reported in Table 4.

Insert Table 4 Here

Step 3 resulted in the following regression equation

difference = $1.4 = 3.80C_1 - 40C_2$ with R^2 equalling .2396.

The computed F statistic was calculated as 1.89.

Burnett and Barr's test statistic has an F distribution with the number of groups minus one and the total sample size minus the number of parameters estimated in the regression model, degrees of freedom. For the present analysis the critical F ratio with 2 and 13 degrees of freedom equalled 3.80 at the .05 level of significance. The probability value associated with the computed F ratio equalled .1984. The results of the analysis therefore indicated that there was insufficient evidence to conclude that the distribucion of posttest scores differed across the three treatment groups.

The rank gain score approach suggested by Burnett and Barr can be used in factorial designs but it is limited to a single covariate and it does

assume the relationship between the pretest and posttest is the same across all groups.

Shirley's general linear models solution. More recently Shirley (1981) suggested a distribution-free alternative to analysis of covariance which was very similar to McSweeney and Porter's (1971) rank transformation approach. Shirley's procedure is based on the work of Bennett(1968) who developed a general model for testing linear models based on ranks. The test statistic suggested by Shirley is computed by taking the ratio of the adjusted sum of squares for the grouping factor and the total mean square error for the unadjusted posttest ranks. This statistic can be computed in the following four steps:

1) For each variable rank the data across groups.
2) Calculate the regression of the ranked dependent variable ranked covariate across all groups and calculate the squared multiple correlation coefficient ($R^2_{RY/RX}$).
3) Calculate the regression of the ranked dependent variable on the ranked covariate and the grouping factor. Determine the multiple squared correlation for this model ($R^2_{RY/RX,G}$).
4) Calculate the adjusted sum of squares for the grouping factor by the product of the sum of squares total on the unadjusted posttest ranks ($\frac{N(N^2-1)}{12}$) and the R^2 increase associated with the grouping factor ($R^2_{RY/RX,G} - R^2_{RY/RX}$).
5) Compute the total mean square on the unadjusted posttest ranks ($\frac{N(N+1)}{12}$) and calculate the ratio of the adjusted sum of squares for the grouping factor to the total mean square for the unadjusted posttest ranks. The ratio reduces to the following $R^2_{increase}(N-1)$.

Applying Shirley's procedure to the data set in Table 1 resulted in the following calculations:

Step 1. Table 2 reports the ranked data set.

Step 2. The computed regression equation of the ranked posttest in the ranked pretest equalled RY = 2.60 + .675RX with $R^2_{RY/RX} = .4556$.

Step 3. The regression equation for the model which included both the ranked pretest and the grouping factor equalled $RY = 3.1794 + .7596RX - 3.271G_1 - .496G_2$ with $R^2_{RY/RX_1G_1G_2} = .5595$.

Step 4. The adjusted sum of squares for the grouping factor equalled: $(.5595 - .4556)(\frac{15(15^2-1)}{12}) = 29.092$.

Step 5. The total mean square equalled 20 and the ratio of the adjusted sum of squares for the grouping factor and the total mean square error equalled 1.455.

Shirley suggests that the test statistic has a chi-square distribution with the number of groups minus one degrees of freedom. For the data set considered here the critical test statistic with two degrees of freedom at a .05 level of significance equalled 5.791. Based on the chi-square distribution the probability value associated with the computed test statistic was equal to .4831. The results of the analysis therefore provided insufficient information to conclude that the distribution of the posttest scores differed across the three groups.

Shirley's procedure can also be used to test the hypothesis regarding the equality of the within group regression slopes across the treatment

groups. The adjusted sums of squares associated with the interaction factor is divided by the total mean square for the unadjusted posttest ranks. Following this procedure the resulting test statistic was computed as 1.512. The probability value associated with this test statistic based on the chi-square distribution equals .4695. The results indicate insufficient evidence to conclude that the relationship between the posttest and pretest differ by groups.

This alternative to analysis of covariance can be used with multiple covariates and with factorial designs.

Summary

For the data set reported by Quade (1967), all five of the nonparametric procedures considered here provided the same conclusions as parametric analysis of covariance. There did not appear to be any treatment effects in that study at the .05 level of significance. The probability value associated with each of the nonparametric test statistics however was less than the p-value associated with the parametric ANCOVA F-ratio. This result may indicate slightly greater statistical power associated with the distribution-free strategies.

The five alternatives to ANCOVA all begin with the same first step, that of ranking each variable across all groups. The subsequent analyses however differ with regard to three important areas. First the nature of the posttest adjustment is different. Quade (1967), Puri and Sen (1969) and Burnett and Barr's (1978) adjustments are based on the relationship between the pre and post rank data for the total sample. McSweeney and Porter (1971) and Shirely (1981) base their adjustment of posttest rank

data on both the within group and total group relationships.

A second general difference between the procedures is the nature of the dependent variable analyzed in computing the test statistic. Although all of the strategies can be reduced to a series of steps, Quade, Puri and Sen as well as Burnett and Barr's procedures can be thought of as a two stage analysis strategies. That is each of these procedures calculates a residual in the first stage of the analysis and in the second stage the residuals are analyzed and a test statistic is computed. McSweeney and Porter and Shirley compute this test statistic in a single stage which is the approach taken in parametric analysis of covariance.

A third difference between the procedures is the reference distribution used with the computed test statistics. Quade, McSweeney and Porter and Burnett and Barr all use the F distribution. Quade and Burnett and Barr have however an additional degree of freedom for the denominator in identifying the critical test statistic than McSweeney and Porter's approach. They both have the same number of degrees of freedom for the numerator. Puri and Sen and Shirley refer their test statistics to the chi-square distribution with the same degrees of freedom (number of groups minus one).

Since the distribution-free procedures differ in several respects there has been some interest in comparing the approaches regarding their Type I error rates and their relative statistical power. Furthermore since these procedures are viewed as alternatives to parametric analysis of covariance these procedures have been compared to parametric ANCOVA when parametric ANCOVA assumptions have been violated as well as when all assumptions have been met. The results of these studies are reviewed in the following section.

First the results regarding Type I error rates are presented and second the analysis of power studies are discussed.

Type I error rates

Parametric assumptions met. Several investigators have examined using monte-carlo procedures to studythe behavior of the distribution-free test statistics when all parametric assumptions weremet. The primary purpose of these investigations was to determine the appropriateness of using an asymptotic reference distribution when the computed test statistic was based on small data sets.

The sampling distributions of Quade's test statistic and the F-ratio computed on rank transformed data were included among the nonparametric strategies studied by McSweeney and Porter (1971). Their primary interest was to compare the precision of these distribution-free analogs to parametric analysis of covariance and three indices of response strategies when parametric assumptions were met. They found a very close fit between the empirically determined sampling distributions of both test statistics and the theoretical F distribution. These results were consistent when the strength of the linear relationship between the pretest and posttest varied (ρ_{xy} = .0,.4,.6,.8) the number of groups varied (3, 5 and 8) and the sample sizes varied (5,8,10). Only the results of a small number of combinations of these three factors were reported, but the results clearly indicated the appropriateness of the central F distribution for these statistics when computed on sample sizes as small as 5 per group.

In a study which focused primarily on the effect of unequal within group regression slopes, Hamilton (1976) provided some data on the sampling distribution of the test statistics suggested by Quade, McSweeney and Porter and

Puri and Sen when all parametric assumptions were met. His results were consistent with McSweeney and Porter's conclusions regarding Quade's test and the rank transformation approach. In addition, Hamilton found that unequal sample sizes in two group studies did not affect the Type I error rate for either procedure. The sampling distribution of Puri and Sen's L_n test statistic however was not found to be consistent with the chi-square distribution when sample sizes were small (10,10). With larger samples (20, 20; 30, 30) and with unequal samples (10, 20; 10, 30 and 20, 30) a close fit was found between the L_n distribution and the chi-square distribution.

Olejnik and Algina (1983) studied the effects of conditional non-normality and heterogeneous conditional variance on parametric and rank transformed ANCOVA but they also reported results of Type I error rates when all parametric assumptions were met. Their results were consistent with the previous findings regarding the sampling distributions of the F ratio based on rank transformed data. In addition, Olejnik and Algina provided evidence indicating that the Type I error rate for the rank transformation approach is not affected by non-normality in the distribution of the covariate.

To support the use of differences in rank data as a nonparametric alternative to analysis of covariance Burnett and Barr (1977) reported the results of a simulation study they had conducted on the sampling distribution of their test statistic under the null case. Their results indicated a very close fit between the sampling distribution of their statistic and the central F-distribution. Their simulation included sample sizes ranging between 4 and 10 units per group and between 2 and 8 groups. Furthermore

the distribution of the pretest data was varied to include: normal, uniform exponential and contaminated normal distributions.

The results of investigations studying the sampling distribution of Shirley's nonparametric alternative to parametric ANCOVA have not as yet been reported in the research literature. A brief simulation study was conducted for the present paper in which the sampling distribution of Shirley's statistic could be examined under the null condition. Data were generated for two groups using the normal functions of the Statistical Analysis System (1979). The pretest (X) and error (E) components were generated separately with $X \sim N(50,1)$ and $E \sim N(0,1)$. The dependent variable Y was generated by combining the pretest and error using the following function:

$$Y = \rho X + E\sqrt{1-\rho^2}$$

where ρ is the Pearson Product Moment correlation relating the dependent variable with the covariate. Shirley had suggested that the test statistic could be computed after either transforming the posttest only data or after transforming both the pretest and posttest data. The present simulation considered both approaches. In addition two levels of relationship were considered in the simulation: a moderate relationship (ρ = .7) and a weak relationships (ρ = .3). Finally five combinations of sample sizes for two groups were considered: (5,5); (10,10); (15,15); (30,30) and (5,10). Each condition was replicated 1000 times. For comparison purposes parametric analysis of covariance and the rank transformed ANCOVA test statistics were also calculated on the generated data sets. For each statistic the proportion of hypotheses rejected were recorded for the .10, .05 and .01 levels of significance.

The results of the simulation study are reported in Table 5.

Insert Table 5 Here

These results indicate that there is a considerable discrepancy between the sampling distributions of Shirley's statistics and the central chi-square distribution. The actual Type I error rate is considerably lower than the nominal Type I error rate. Furthermore the stronger the relationship between the pretest and posttest the greater the discrepancy. The sampling distributions for both the parametric and rank transformed ANCOVA test statistics on the other hand show a very close fit with the central F distribution. Based on these findings it can be concluded that Shirley's procedure leads to a conservative hypothesis test and as a result it does not provide a reasonable alternative to parametric analysis of covariance.

Nonlinearity. Parametric analysis of convariance assumes that the relationship between the pretest and posttest is linear. In many research settings this may be an unrealistic assumption. Nonlinearity may occur for example in situations where floor or ceiling effects are present. Since the distribution-free strategies only require a monotonic increasing or decreasing relationship, they potentially provide a very useful alternative to parametric analysis of covariance. Only one investigation was identified which examined the sampling distribution of a nonparametric alternative to ANCOVA when the relationship between the pretest and posttest was nonlinear. Porter and McSweeney (1974) considered the rank transformation approach when the relationship between the measures was exponential. They attempted to simulate a situation where a ceiling effect was present. The result of

their analysis indicated that Type I error rates for both parametric and rank transform analysis of covariance was unaffected by the lack of linearity between the measures. These conclusions were consistent across three levels of eta squared (.4, .6 and .8) for sample sizes of 5, 8 and 10 for a grouping factor having 3, 5 and 8 levels. All combinations of these three factors were not included but the consistency of the reported results clearly indicated that a factorial combination of factors was not necessary.

Non-normality. Interest in nonparametric analysis strategies developed as a result of the belief that the normality assumption made with parametric procedures was too restrictive for many research settings. In parametric analysis of covariance the normality assumption is made regarding the conditional distributions of the posttest measure. Several investigations have been conducted examining the effect of conditional non-normality on Type I error rates for both the distribution-free as well as the parametric test statistics.

Conover and Iman (1982) simulated situations where the conditional distributions were lognormal, exponential, uniform and Cauchy. In a comparison of Type I error rates for parametric ANCOVA, rank transform ANCOVA and Quade's test statistic, they found the Type I error rates for the latter two strategies to be very similar to the nominal (.05) significance level. The parametric procedure however underestimated the nominal significance level for the lognormal and Cauchy conditional distributions but provided appropriate Type I error rates when the conditional distributions were exponential and uniform.

In a study simulating less extreme departures from conditional normality involving both a normal and non-normal covariate distribution, Olejnik and Algina (1983) found the Type I error rate for parametric ANCOVA to be similar to the nominal alpha level. The empirically derived Type I error rate for the rank transform approach however overestimated the nominal

singificance level when the sample sizes were small and alpha equalled .01. With moderate sample sizes the observed Type I error rate for both strategies did not deviate greatly from the nominal significance levels. These results were consistent across equal and unequal sample sizes and for both a weak (ρ = .3) and a moderate (ρ = .7) relationship between the pretest and posttest.

Finally Abunnaja (1983) studied conditional distributions which were uniform, double exponential and translated exponential with equal and unequal sample sizes for two and three group designs. The results of his analysis indicated that the Type I error rates for parametric ANCOVA, rank transformation, Quade and Burnett and Barr's approaches were not seriously affected by the conditional non-normality. Only Puri and Sen's approach overestimated the nominal significance level in two group designs with sample sizes of 10 units per group. For larger samples (30,30; 10,30; 20,30) and for three group designs the observed Type I error rate did not deviate significantly from the nominal significance level.

Heterogeneous regression slopes. The effect of violating the assumption of equal within-group regression slopes on the Type I error rate for parametric analysis of covariance has been the subject of several research investigations. The results of these studies have indicated that parametric ANCOVA is robust to violations of this assumption when sample sizes are equal. If sample sizes differ however, the actual Type I error rate can either over or underestimate the nominal significance level depending on the relationship between slope discrepancies and sample size difference. (Glass, Peckham and Sanders, 1972; Hsu, 1983). Violation of this assumption have also received some attention with regard to Type I error rates for the test statistics of the distribution-free ANCOVA procedures.

Hamilton (1976) investigated the sampling distributions of the test statistics for parametric ANCOVA, Quades test, rank transformation and Puri and Sen's procedure. In his study twenty different combinations of discrepant within group regression slopes were considered for two group designs having either equal or unequal sample sizes. For designs having equal n, Hamilton found that with the exception of Puri and Sen's procedure, the analysis strategies were robust to discrepant regression slopes. With moderate sample sizes (20, 20 or 30, 30) Puri and Sen's test statistic also followed its chi-square reference distribution but with small samples (10,10) an inadequate fit was found. In situations where the sample sizes differed (10, 20; 10, 30; 20, 30) none of the analysis strategies studied fit their respective reference distribution.

It has been pointed out however that in generating his data, Hamilton unintentionally also generated data which violated the assumption of equal conditional error variance between groups.(Rogosa, 1980) Abunnaja (1983) has recently investigated the problem and has generated data in which only the relationship between the measures differed and not the between group error variances. In addition to the four procedures considered by Hamilton, Abunnaja also included Burnett and Barr's test statistic. Four combinations of slope discrepancies two equal and three unequal sample size combinations were included factorially for both two group and three group designs. The results of his analysis were consistent with Hamilton's conclusions with two exceptions. First Abunnaja's results indicated that small discrepances between regression slopes did not seriously affect the Type I error rate of parametric ANCOVA if the sample size differences were not great. Second in two group designs, Puri and Sen's test statistic appeared to follow the chi square distribution

even with small samples (10,10). In three group designs the Type I error rate deviated considerably from the nominal significance level when samples were small.

The nonparametric alternatives suggested by Quade, McSweeney and Porter and Puri and Sen did not appear to do any better in terms of their robustness than parametric ANCOVA. In fact for the three group designs the discrepancies between the actual and nominal significance levels were greater for the distribution-free procedures than for parametric ANCOVA.

In balanced designs, Burnett and Barr's statistic had Type I error rates very similar to the nominal significance level in both two and three group designs. However with unequal sample sizes and discrepant regression slopes the Type I error rate either over or underestimated the nominal alpha.

Abunnaja also considered the effect of violating simultaneously both of assumptions of conditional normality and equal slopes for the same situations described above. The results of these analyses indicated that when sample sizes were equal parametric ANCOVA was robust to the dual violation of conditional nonnormality and homogeneous slopes. The nonparametric alternatives were also robust to the dual violation except when the conditional distributions were translated exponential. For the combination of unequal slopes and a translated exponential conditional distributions the actual Type I error rates overestimated the nominal significance levels. All of the analysis procedures were more sensitive to the dual violation of assumptions when the sample sizes were unequal. Generally, the actual Type I error rate for all of the procedures overestimated the nominal significance level when the discrepancy in slopes and difference in sample sizes were large. An exception to this

generalization was provided by Burnett and Barr's statistic which underestimated the nominal significance level when sample sizes were (30,10) or (30,20,10).

Heterogeneous conditional variance. The assumption of homogeneous conditional variances can be violated in two ways. In one case the conditional variances between groups may be the same but the within group conditional variances may vary. Second, the within group conditional variances could be constant but differ between groups. Nonparametric analysis strategies are sensitive to between group variance discrepancies. As a result the effect of heterogeneous conditional variance of the within group type have been studied for the nonparametric alternatives to ANCOVA. Shields (1979) has shown that parametric analysis of covariance is robust to heterogeneous conditional within group variances for both equal and unequal sample size designs.

Only the McSweeney and Porter's rank transformation procedure has been considered in situations where the within group conditional variances differed. Olejnik and Algina (1983) considered situations where the conditional variances increased with values of the pretest. Their investigation was limited to two group studies but included both equal and unequal sample sizes, covariates having normal and non-normal distributions and both weak and moderate relationships between the covariate and the posttest. The empirical Type I error rate for the rank transform F ratio was very similar to the nominal significance level for all conditions studied. They also considered the dual violation of conditional normality and homoscedasticity. The results indicated that the Type I error rate overestimated the nominal significance

level was .01. For other conditions studied there was close agreement between the observed Type I error rate and the nominal alpha level.

Summary

The simulation studies which have investigated Type I error rates for distribution-free alternatives to parametric analysis of covariance have indicated that these procedures provide appropriate Type I error rates for those situations where parametric ANCOVA is robust to violations of data assumptions. In situations where parametric ANCOVA is not robust (e.g. unequal slopes with unequal sample sizes) the nonparametric are also inappropriate. Furthermore simulation results presented in the present paper indicates that Shirley's application of Bennett's general linear model for ranked data leads to a conservative hypothesis test and thus should not be considered an appropriate alternative to parametric analysis of covariance. Simulation results also provide evidence that for small data sets, Puri and Sen's test statistic does not approximate the chi-square distribution. For small samples the exact probability of the test statistic should be computed. Finally Burnett and Barr's test gains in rank scores was shown to provide Type I errors which over or underestimates the nominal significance levels when sample sizes are unequal and the within group regression slopes differ.

Statistical Power

All assumptions met. When all parametric assumptions have been met, a comparison of statistical power estimates is of interest to determine the degree to which precision is lost when a distribution-free procedure is chosen rather than on appropriate parametric approach. Quade calculated the asymptotic relative efficiency of his statistic relative to parametric ANCOVA

when all parametric assumptions were met and found it to equal:

$$.866 \leq \frac{3(1-\rho_{xy}^2)}{\Pi(1-\rho_s^2)} \leq .955.$$

Thus very little efficiency is lost in using the distribution-free statistic suggested by Quade when assumptions are met. When assumptions have not been met the nonparametric approaches may be considerably more sensitive to treatment effects than the parametric strategies. Comparisons of power estimates between parametric ANCOVA and the distribution-free approaches has received considerable attention. When all parametric assumptions are met the results of these studies have indicated that very little power is lost when a nonparametric procedure is used.

McSweeney and Porter (1971) in considering an effect size of a standard deviation found most power differences between parametric ANCOVA, the rank transform ANCOVA and Quade's procedure to be less than 5% favoring the parametric approach. Slightly greater power differences (less than 10%) were observed favoring parametric ANCOVA when there was a strong linear relationship between the measures ($\rho_{xy} = .8$).

For two group designs which differed by a half standard deviation, Hamilton (1976) found parametric ANCOVA to provide less than 5% greater power than the rank transform, Quade's procedure and Puri and Sen's approach. With moderate sample sizes (20,20; 30,30) Puri and Sen's approach provided slightly greater power than either the rank transform or Quade's statistic which were almost identical.

Olejnik and Algina (1984) developed empirical power curves for the parametric and rank transform ANCOVA. Power estimates for the two procedures were generated for effect sizes of .1 to .8 standard deviation units

increasing in .1 units for a two group design with 20 subjects per group. The results indicated that for a weak relationship between the measure (ρ_{xy}=.3) the power advantage for parametric ANCOVA was less than 5%. With a moderate relationship (ρ_{xy}=.7) between the measures the power advantage for parametric ranged between .5 and 8.5 percent.

Finally power estimates reported by Abunnaja (1983) reflected similar patterns as those cited above for the parametric, rank transform, Quade and Puri and Sen procedures. In addition, the statistical power estimates reported for Burnett and Barr's procedure were considerably lower than any of the other distribution-free procedures studied. These results were not surprising since the data generated by Abunnaja had a .5 relationship between the pretest and posttest measures. The gains in rank scores approach assumes a perfect relationship between the measures and the lower power estimates reflect the magnitude of the deviation from the assumed relationship.

Nonlinearity. When the population relationship between the pretest and posttest was exponential, Porter and McSweeney (1974) found substantial power advantages associated with the rank transformation approach over linear parametric ANCOVA. They considered an effect size of a half standard deviation under various combinations of a) strengths of curvilinear relationship (η^2 = 14,16 or .8) b) a grouping factor having 3, 5 or 8 levels and c) sample sizes of 5, 8 or 10 per group. The rank transformation approach was always more powerful linear parametric ANCOVA with power differences as great as 13% for η^2=.8, n=8 when there were 8 levels of the grouping factor. The researchers did not compare power estimates when a curvilinear ANCOVA model was used to analyze the generated data.

Non-normality. Asymptotic relative efficiency studies comparing parametric and nonparametric analysis strategies have shown that when the parametric assumption of normality has been violated, nonparametric procedures may require significantly fewer subjects than parametric alternatives to achieve equal power (Noether, 1967). It might be expected then that the distribution-free ANCOVA procedures would provide greater power than parametric ANCOVA when the conditional distributions were non-normal. Results consistent with those expectations were reported by Conover and Iman (1982). In a statistical power demonstration involving four groups with an equal n of 10, and an effect size of a half standard deviation, Conover and Iman reported power differences as great as 57% favoring the Quade's and the rank transform approaches when the conditional distributions were lognormal. Smaller power differences were observed still favoring the distribution-free approaches when the conditional distributions were exponential (25%) and Cauchy (30%). The two nonparametric approaches provided almost identical power estimates. When the conditional distribution was uniform however a 7% power advantage favoring the parametric approach was observed.

In two group designs where the relationship between the pretest and posttest was .5 and an effect size of a half standard deviation was studied, Abunnaja (1983) found less dramatic power differences than those reported by Conover and Iman. Similar power differences were obtained in equal and unequal sample size situations. With a uniform conditional distribution a 3 to 5 percent power advantage favoring the parametric approach was reported. When the conditional distributions were double exponential the distribution-free procedures provided a 3 to 6 percent power advantage. And when the conditional distributions were translated exponential the nonparametric

procedures had a 6 and 16 percent power advantage over parametric ANCOVA. The distribution-free procedures suggested by Quade, McSweeney and Porter and Puri and Sen all provided similar power estimates.

The studies conducted by both Conover and Iman and Abunnaja compared power estimates for the distribution-free and parametric ANCOVA when a moderately weak relationship existed between the pretest and posttest. Olejnik and Algina (1984) generated data in which both a weak (ρ=.3) and a moderately strong (ρ=.7) relationship existed between the measures and compared power curves for the rank transform with parametric ANCOVA. Their results indicated that parametric ANCOVA power was not seriously affected by small departures from conditional normality. Furthermore when the relationship between the pretest and posttest was moderately strong (ρ=.7) almost no difference in statistical power was observed between the procedures except when the conditional distributions were platykurtic or highly skewed and moderately leptokurtic. For the former conditional distribution the parametric approach provided a power advantage of between .6 and 14.3 percent. On the other hand for the highly skewed, moderately leptokurtic conditional distribution the rank transformation approach provided a power advantage between 1.3 and 8.9 percent. For the weak relationship (ρ=.3) between the measures, power advantages associated with the rank transform approach similar to those reported by Abunnaja (1983) were observed.

Heterogeneous regression slopes. As discussed earlier, parametric and distribution-free procedures are robust to violations of the assumption of equal within-group regression slopes when samples sizes are equal. Hamilton (1976) considered these situations and compared statistical power estimates for procedures suggested by Quade, McSweeney and Porter and Puri and Sen

with parametric ANCOVA for balanced two group designs having an effect size of a half standard deviation. Sample size combinations of 10,10; 20,20 and 30,30 provided similar conclusions. Across all 20 combinations of discrepant regression coefficients, parametric ANCOVA consistently provided a small power advantage over the nonparametric alternatives. Differences in power were generally around 2 percent and were never greather than 5 percent. All of the nonparametric procedures provided similar power estimates.

Abunnaja (1983) calculated power estimates for the same procedures as Hamilton plus the procedure suggested by Burnett and Barr (1977). He also considered situations where sample sizes differed. For those situations where appropriate Type I error rates were observed, Abunnaja's results were consistent with Hamilton's conclusion. Parametric ANCOVA was slightly more powerful than the distribution-free alternatives including Burnett and Barr's statistic.

Statistical power estimates for situations where both assumptions of conditional normality and homogeneity of regression slopes were violated was also studied by Abunnaja. In balanced two group designs, small power advantages for parametric ANCOVA were obtained when the conditional distributions were unifrom across all discrepant slope combinations. When the conditional distributions were double exponential the opposite conclusions were drawn, the nonparametric procedures had a slight advantage. Type I error rates for the distribution-free procedures overestimated the nominal significance level when the conditional distributions were translated exponential and therefore a comparison of power estimates is not meaningful. However parametric ANCOVA had empirical Type I error rates similar to the

nominal levels and the results indicated that for situations where the conditional distributions were translated exponential and within group regression slopes differed greater power (5%) was achieved than when the slopes were unequal and the conditional distribution was normal. Thus parametric ANCOVA gained statistical power under this combination of dual violations of assumptions.

For situations where the sample sizes differed and the dual violations of assumptions involved small differences in regression slopes parametric ANCOVA provided slightly greater power when the conditional distributions were uniform. Slight power advantages favoring the distribution-free procedures were obtained when the conditional distributions were either double exponential or translated exponential. If the dual violations of assumptions included large discrepancies in regression slopes then actual Type I error rates exceeded the nominal significance levels for either the parametric or nonparametric procedures and thus comparisons of statistical power for those situations was meaningless.

Heterogeneous error variance. Olejnik and Algina (1983 , 1984) considered situations where the conditional variances within groups increased as the values of the pretest increased but the increasing conditional variances were the same between groups. For this set of conditions they found almost no differences in the power estimates generated for the parametric and rank transform ANCOVA. These results were consistent across several rates of increasing conditional variances and for situations where η^2 equalled .3 and .7.

The effect of dual violations of conditional normality and homogeneous within-group error variance was also studied by Olejnik and Algina (1984) When both assumptions were violated considerable power differences were

observed depending on the shape of the conditional distributions and the strength of the relationship between the pretest and posttest. For η^2 equalling .7 and a platykurtic conditional distribution, the parametric ANCOVA provided greater statistical power ranging between 1.4 and 9.1 percent, while with a highly skewed, moderately leptokurtic conditional distribution the rank transform procedure provided a power advantage ranging between .8 and 13.9 percent. Less exteme departures from conditional normality had little effect on the power difference between the procedures. For η^2 equalling .3 very small differences in power were observed for the platykurtic conditional distribution. Power differences favoring the rank transformation approach began with symmetric slightly leptokurtic conditional distributions and increased as skewness was added to the conditional distributions. With the highly skewed moderately leptokurtic conditional distributions the rank transform ANCOVA provided power advantages ranging between 2.6 and 27.5 percent over the parametric ANCOVA.

Summary

The results of the simulation studies on statistical power of the distribution-free alternatives to analysis of covariance indicate that the procedures suggested by Quade, McSweeney and Porter and Puri and Sen all provide similar power for situations where parametric assumptions were met and where assumptions were violated. The study which considered the Burnett and Barr's statistic considered only situations where the relationship between the pretest and posttest was weak and as a result this approach was consistently shown to provide lower power estimates than the other nonparametric procedures. Since Shirley's nonparametric alternative was shown to have actual Type I error lower than the nominal significance levels estimates of statistical

power for this procedure would have been meaningless and therefore have not been considered.

In comparison with parametric analysis of covariance the distribution-free alternatives provide the greatest power advantages when the linearity assumption is violated and when at least two parametric assumptions were violated. The results regarding nonlinearity however were based on a comparison of a nonparametric procedure with a linear ANCOVA model. If an appropriate curvilinear ANCOVA model had been considered the results may have been different. With respect to dual violations of assumptions the power advantages associated with distribution-free procedures appears to be related to the strength of the linear relationship between the pretest and posttest. When a weak relationship exists between the measures a considerable power advantage favoring the nonparametric approaches occurred. As the relationship between the measures increases the power differences are reduced.

Finally when all parametric assumptions had been met very little statistical power is lost when a distribution-free procedure is used rather than parametric analysis of covariance.

Conclusions

Among the conclusions drawn from their review of literature, Glass, Peckham and Sanders (1972) asserted that given the robustness of parametric procedures to violations of data assumptions the need for nonparametric procedures is marginal. In a reaction to that conclusion Blair (1981) argued that nonparametric procedures provide greater statistical power than parametric strategies when data assumptions are violated and therefore should

be the preferred analysis strategy. The present paper has reviewed the empirical research evidence of Type I error rates and statistical power for several distribution-free alternatives to parametric analysis of covariance. The results of that review indicates that for those situations where parametric ANCOVA is robust to data assumptions the Type I error role for the nonparametric alternatives also closely reflects the nominal significance level. But when parametric ANCOVA is not robust to assumption violations neither were the nonparametric alternatives. With regard to statistical power the results indicated that unless extreme violations of assumptions occurred or unless more than one assumption was violated in a data set, the distribution-free strategies did not provide significant power advantages over the parametric analysis of covariance. Furthermore, the differences in statistical power appeared to be moderated by the strength of the linear relationship between the pretest and posttest. The stronger the relationship between the measures the smaller the power advantage for the nonparametric procedures. Since for most actual research studies extreme violations of assumptions are unlikely and in general covariates are chosen because of their strong relationship with the dependent variable, there does not appear to be convincing evidence to support the recommendation of using distribution-free procedures rather than parametric analysis of covariance. If however extreme violations of assumption do occur and the relationship betwen the covariate and posttest is weak then considerable power advantage can be achieved by using the distribution-free procedures.

References

Abunnaja, S. S. (1983) The robustness of parametric and nonparametric analysis of covariance against unequal regression slopes and non-normality. Paper presented at the meeting of the American Educational Research Association in Montreal, Canada.

Bennett, B. M. (1968) Rank-order tests of linear hypotheses. J. Statist. Soc B, 30, 483-489.

Blair, C. (1981) A reaction to "Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance." Review of Educational Research, 51, 499-508.

Blair, R. C. & Higgins, T. J. (1980) A comparison of the power of Wilcoxon's rank-sum statistic to that of Student's t statistic under various non-normal distributions. Journal of Educational Statistics, 5, 309-335.

Burnett, T.D. & Barr, D. R. (1977) A nonparametric analogy of analysis of covariance. Educational and Psychological Measurement, 37, 341-348.

Campbell, D. T. & Stanley, J. C. (1963) Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Conover, W. J. & Iman, R. L. (1982) Analysis of covariance using the rank transformation. Biometrics, 38, 715-724.

Elashoff, J. D. (1969) Analysis of Covariance: A delicate instrument. American Educational Research Journal, 6, 383-401.

Glass, G. U, Peckham, P. D. & Sanders, J. R. (1972) Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research, 42, 237-288.

Hamilton, B. L. (1976) A monte-carlo test of robustness of parametric and nonparametric analysis of covariance against unequal regression slopes. Journal of the American Statistical Association, 71, p. 864-869.

Hsu, T. (1983) The robustness of ANCOVA to the violation of various assumptions: A review of recent studies. Paper presented at the meetings of the American Educational Research Association in Montreal, Canada.

McSweeney, M. & Porter, A. C. (1971) Small sample properties of nonparametric index of response and rank analysis of covariance. Occasional paper No. 16, Office of Research Consultation. Michigan State University, East Lansing, Michigan.

Noether, G. E. (1967) Elements of Nonparametric Statistics. John Wiley, New York.

Olejnik, S. F. & Algina, J. (1983) Parametric ANCOVA vs. rank transform ANCOVA when assumptions of conditional normality and homoscedasticity are violated. Paper presented at the meeting of the American Educational Research Association in Montreal, Canada.

Olejnik, S. F., Algina, J. and Abdel-Fattah, A. F. (1984) An analysis of statistical power curves for parametric ANCOVA and rank transform ANCOVA. Paper presented at the meeting of the American Educational Research Association in New Orleans, Louisiana.

Porter, A. C. & McSweeney, M. (1974) Comparison of three common experimental designs to improve statistical power when data violate parametric assumptions. Paper presented at the meeting of the American Educational Research Association.

Puri, M. L. & Sen, P. K. (1969) Analysis of covariance based on general rank scores. Annals of Mathematical Statistics, 40, 610-618.

Quade, O. (1967) Rank analysis of covariance. J. Amer. Statis. Assoc., 62, 1187-1200.

Rogosa, D. (1980) Comparing nonparallel regression lines. Psychological Bulletin, 88, 307-321.

Shields, J. L. (1978) An empirical investigation of the effect of heteroscedasticity and heterogeneity of variance on the analysis of covariance and the Johnson-Neyman technique. Technical paper No. 292, U. S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia.

Shirley, E. A. (1981) A distribution-free method for analysis of covariance based on ranked data. Journal of Applied Statistics, 30, p. 158-162.

Table 1

Quade's (1967) artificial raw data for a three group pretest-posttest design

Group	Posttest	Pretest
G	Y	X
1	16	26
1	60	10
1	82	42
1	126	49
1	137	55
2	44	21
2	67	28
2	87	5
2	100	12
2	142	58
3	17	1
3	28	19
3	105	41
7	149	48
3	160	35

Table 2

Ranked pretest and posttest data across the three participating groups

Group	Posttest	Pretest
G	RY	RX
1	1	7
1	5	3
1	7	11
1	11	13
1	12	14
2	4	6
2	6	8
2	8	2
2	9	4
2	13	15
3	2	1
3	3	5
3	10	10
3	14	12
3	15	9

Table 3

Individual residual scores based on the regression of the ranked posttest on the ranked pretest data for the total sample

Group	RY-RY
1	-6.325
1	.375
1	-3.025
1	-.375
1	-.050
2	-2.650
2	-2.000
2	4.050
2	3.700
2	.275
3	-1.275
3	-2.975
3	.650
3	3.300
3	6.325

Table 4

Difference between ranked posttest and ranked pretest data

Group	RY-RX
1	-6
1	2
1	-4
1	-2
1	-2
2	-2
2	-2
2	6
2	5
2	-2
3	1
3	-2
3	0
3	2
3	6

Table 5

Observed proportion of Type I errors for four ANCOVA procedures under varying sample sizes, strength of relationship and nominal significance levels

n_1	n_2	ρ	α	ANCOVA[a]	χ^{2}[b]	χ^{2}[c]	RANCOVA[d]
5	5	.3	.01	.013	.002	.004	.017
			.05	.053	.034	.037	.050
			.10	.100	.089	.083	.103
		.7	.01	.012	.000	.000	.009
			.05	.046	.009	.011	.044
			.10	.088	.033	.030	.088
10	10	.3	.01	.007	.005	.005	.010
			.05	.036	.038	.035	.044
			.10	.082	.068	.067	.079
		.7	.01	.012	.001	.001	.013
			.05	.047	.013	.012	.054
			.10	.097	.032	.036	.111
15	15	.3	.01	.009	.002	.003	.007
			.05	.049	.040	.036	.050
			.10	.103	.099	.097	.107
		.7	.01	.012	.004	.004	.010
			.05	.046	.008	.012	.049
			.10	.094	.022	.026	.099
30	30	.3	.01	.010	.009	.009	.010
			.05	.044	.043	.046	.051
			.10	.095	.076	.078	.090
		.7	.01	.009	.000	.000	.010
			.05	.039	.009	.009	.047
			.10	.087	.023	.023	.090
5	10	.3	.01	.009	.002	.002	.013
			.05	.060	.044	.043	.063
			.10	.105	.099	.095	.103
		.7	.01	.011	.000	.000	.009
			.05	.042	.009	.010	.054
			.10	.092	.027	.034	.104

[a]Parametric analysis of covariance

[b]Bennett's procedure with dependent variable alone transformed with ranks.

[c]Bennett's procedure with both the dependent variable and the covariate transformed with ranks.

[d]McSweeney and Porter's procedure with both the dependent variable and covariate transformed with ranks.